**3Б**

**УДК 004.822**

***О.В. Бісікало, Г.О. Кириленко***
Вінницький національний технічний університет, Україна
Україна, 21021, м. Вінниця, вул. Хмельницьке шосе, 95

# Застосування нечіткої логіки для визначення сили зв’язку між мовними образами

***O.V. Bisikalo, H.O. Kyrylenko***
*Vinnytsia National Technical University, Ukraine*
*Ukraine, 21021, c. Vinnytsia, Khmelnytske shoes st., 95*

# *Using of fuzzy logic for determining of connection strength between language images*

***О.В. Бисикало, А.А. Кириленко***
Винницкий национальный технический университет, Украина
Украина, 21021, г. Винница, ул. Хмельницкое шоссе, 95

# Применение нечеткой логики для определения силы связи между языковыми образами

У статті розглянуто спосіб побудови таблиці зв’язків між мовними образами з метою створення онтологій тексту, а також запропоновано використання методів нечіткої логіки для визначення сили зв’язку між мовними образами.
**Ключові слова:** мовний образ, синтаксичний зв’язок, нечітка логіка, функція належності.

The article considers method of building of table with connections between language images for text ontologies creation, also fuzzy logic methods usage are proposed for determining of strength value between language images.
**Key words:** language image, syntactic connection, fuzzy logic, membership function.

В статье рассмотрен способ построения таблицы связей между языковыми образами з целью создания онтологий текста, а также предложено использование методов нечеткой логики для определения силы связи между языковыми образами.
**Ключевые слова:** языковой образ, синтаксическая связь, нечеткая логика, функция принадлежности.

## Вступ

Комп’ютерне моделювання мовленнєвої діяльності людини є однією з базових проблем в галузі побудови інтелектуальних систем. Напрями його застосування найрізноманітніші – це технологія для машинного перекладу, діалогові системи, визначення авторства, автореферування, системи інформаційного пошуку тощо.

При побудові людино-машинних інтелектуальних систем необхідним є представлення тексту в такому вигляді, щоб комп’ютер міг ефективно обробити відповідну природно-мовну інформацію. Найзручнішим способом отримання знань є аналіз текстової інформації. Саме тому сьогодні широко застосовуються онтології – формальні представлення знань певної предметної області. Вони використовуються в предметних областях медицини, біоінформатики, семантичної павутини тощо. Особливої актуальності набуває сьогодні задача автоматизації побудови онтологій.

Побудова онтологій вимагає значних витрат часу роботи людини-експерта. Іншою значною проблемою є суб'єктивність, яку вносить кожний автор, а тому кінцева онтологія потребує знаходження «спільного знаменника». Отже, маємо важливу задачу автоматизації процесу отримання знань з тексту з метою побудови онтології. Розв'язок цієї задачі значно зменшить витрати на створення онтологій, а роботу експерта можна буде застосовувати лише для оцінки кінцевих результатів, оскільки в даній задачі повністю виключити людський фактор практично неможливо.

Відомі системи обробки текстової інформації базуються зазвичай на автоматичному визначенні ключових слів, що ставляться у відповідність до значимих понять предметної області. При цьому зв'язки між такими поняттями, як правило, вносяться у онтологію вручну [1]. В роботі [2] запропоновано застосувати модель образного мислення людини з метою автоматизації отримання асоціативних зв'язків різних типів між мовними образами. Останніми вважають множини однокореневих слів, які характеризують окремий образ з нескінченної множини $I = \{i_1, i_2, ..., i_n, ...\}$ . Запропонований підхід забезпечує морфемну класифікацію та гніздовий принцип організації словника мовних образів. Розглянуті в роботі [3] нечітке відношення і простір сенсу образних конструкцій забезпечують формальну основу для образної індексації природномовного контенту.

**Мета і задачі.** Робота націлена на створення методу автоматизованої побудови онтологій, який базується на образній індексації електронного контенту та забезпечує використання методів нечіткої логіки для визначення сили зв'язку між мовними образами. Для досягнення мети необхідно розв'язати такі задачі:

– побудова таблиці семантичних зв'язків між мовними образами у тексті;

– визначення сили зв'язку між мовними образами та побудова онтологій тексту на основі цієї інформації.

## Формування таблиці зв'язків між мовними образами

Запропоноване поняття мовного образу спирається на корінь слова, оскільки саме коренева послідовність символів природним шляхом об'єднує словоформи різних частин мови. Із самостійних частин мови було обрано найбільш значимі – іменник, прикметник, дієслово та прислівник. З них будується словник мовних образів $I' \subset I$ у вигляді п'ятірки концептів $I' = \langle OQ, O, N, M, MQ \rangle$, де $OQ$ – якість об'єкта, $O$ – об'єкт, $N$ – поняття, $M$ – метод, $OQ$ – якість методу. Оскільки обрані концепти можна поставити у відповідність членам речення (означення, додаток, підмет, присудок і обставина), з'являється можливість фіксувати синтаксичні зв'язки як основу узагальнення онтологічних.

Враховуючи, що кінцева система підтримки онтологій має обробляти фахові тексти з предметної області, для побудови моделі образної індексації електронного контенту потрібно послідовно розв'язати такі задачі: отримати текстову інформацію з електронного контенту, визначити фахову придатність тексту, виокремити речення тексту, для кожного речення поставити у відповідність словам речення мовні образи та побудувати граф синтаксичних зв'язків між мовними образами, об'єднати окремі графи у загальний для всього тексту, побудувати онтологію із загального графа на основі вагових параметрів зв'язків.

Для прикладу взято текст про базу даних, фрагмент якого наведено нижче:

*«База даних (БД) – це організована структура, призначена для зберігання інформації: даних і методів, за допомогою яких відбувається взаємодія з іншими програмно-*

*апаратними комплексами. Системи управління базами даних (СУБД) – це комплекс програмних засобів, призначених для створення структури, наповнення її змістом, редагування змісту та візуалізації інформації».*

Отже, процес побудови онтологій буде складатись з декількох кроків. На першому кроці відкидаємо в тексті всі розділові знаки, а також слова, які не несуть вагомого змісту згідно з [2], [3]. У результаті отримуємо лише послідовність тих слів, з яких побудуємо словник мовних образів:

*«база даних організована структура призначена зберігання інформації даних методів допомогою відбувається взаємодія програмно апаратними комплексами системи управління базами даних комплекс програмних засобів призначених створення структури наповнення змістом редагування змісту візуалізації інформації».*

Далі будуємо таблицю, де по горизонталі і вертикалі розташовано відсортовані за алфавітом мовні образи. У комірку на перетині кожного окремого рядка і стовпця додаємо 1 у тому випадку, коли в реченні є синтаксичний зв'язок між цими двома словами.

Для зручного представлення таблицю можна відсортувати за кількістю зв'язків, що зустрічаються, і розмістити найбільш вагомі з них у правому верхньому куті. Також виділимо різні за вагою зв'язки різними за інтенсивністю кольорами. Темніший колір – сильніший зв'язок. Приклад подібної таблиці представлено на рис. 1.

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | 1 - 3 | | | | | | | | | | | |
| 2 | | 4 - 10 | | | дані, вид | модель | база | таблиця | елемент | зв'язок | знак | сегмент | інформа |
| 3 | | > 10 | | | даний | модельн | базовий | табличн | елемент | звязани | признач | сегмент | інформа |
| 4 | | | | | давати | моделюі | базувати | | | звязува | призначати | | інформу |
| 5 | | | | | дано | | базово | | елемент | звязано | значно | | інформа |
| 6 | модель | модельн | моделювати | | 8 | | 3 | | 1 | | 1 | | 2 |
| 7 | база | базовий | базуват | базово | 17 | | | | | | 1 | | |
| 8 | знак | признач | признач | значно | | 1 | 1 | | 5 | 1 | | | |
| 9 | елемент | елементарний | | елемент | 7 | | | 1 | | 2 | | 1 | |
| 10 | дані, вид | даний | давати | дано | 2 | | | | | 1 | | | |
| 11 | зв'язок | звязани | звязува | звязано | 2 | 3 | | 1 | 2 | | 2 | | |
| 12 | склад | складен | скалада | складено | | 4 | | 3 | | | | 1 | |
| 13 | буття | | бути, відбуватись | | | | 3 | 1 | | | | | |
| 14 | сегмент | сегментний | | | | | | | | | | | |
| 15 | таблиця | табличний | | | | | | | | 2 | | | |
| 16 | порядок | упорядк | упорядк | упорядк | 1 | | | | 1 | | | | |
| 17 | система | системн | системт | системн | 1 | | | | | | | | |
| 18 | екземпляр | | | | | | | | | 1 | 1 | 4 | |

Рисунок 1 – Таблиця зв'язків між мовними образами

Для даного експериментального прикладу вибрано попередньо оброблений текст про бази даних. За результатами експертним шляхом було визначено, що для побудови онтологій мають значення лише зв'язки, які зустрічаються в тексті більше 4 разів. А ті, що зустрічаються більше 10 разів, мають найбільше значення. Так, наприклад, зв'язок між словами «база» і «дані» зустрічається 17 разів.

Для інших текстів кількість повторення зв'язків, яку можна вважати вагомою, може відрізнятись. Якщо проаналізувати багато текстів, то можна визначити усереднені значення слабких, сильних і дуже сильних зв'язків (класифікацію можна розширювати). Тому пропонується використовувати методи нечіткої логіки для вирішення задачі класифікації сили зв'язку.

**Застосування нечіткої логіки для визначення сили зв'язку.** При описі об'єктів і явищ за допомогою нечітких множин використовується поняття лінгвістичної змінної. Лінгвістичною змінною називається набір <$\beta$, $T$, $X$, $G$, $M$>, де:

1) $\beta$ – найменування лінгвістичної змінної;

2) $T$ – множина її значень (терм-множина), що представляють собою імена нечітких змінних, області визначення, кожної з яких є множина X;

3) $G$ – синтаксична процедура, що дозволяє оперувати елементами терм-множини T;

4) $M$ – семантична процедура, що дозволяє перетворити кожне нове значення лінгвістичної змінної, утвореної процедурою G, у нечітку змінну, тобто сформувати відповідну нечітку множину [4].

Розглянемо тепер поняття сили зв'язку. У даному випадку створюємо лінгвістичну змінну «сила зв'язку». Зв'язок може бути слабким, сильним та дуже сильним. Зв'язки можуть зустрічатись в тексті з частотою від 0 до 100%. Формалізація такого опису може бути проведена за допомогою лінгвістичної змінної $< \beta, T, X, G, M >$, де:

1) $\beta$ – «сила зв'язку»;

2) $T$ – {«слабкий зв'язок», «сильний зв'язок», «дуже сильний зв'язок»} ($X \in [0,100]$);

3) $G$ – процедура утворення нових термів за допомогою зв'язувань «і», «або» і модифікаторів типу «дуже», «не», «злегка» та ін.

4) $M$ – процедура завдання на X = [0, 100] нечітких підмножин A1= «слабкий зв'язок», A2 = «сильний зв'язок», A3 = «дуже сильний зв'язок».

Найбільше у нечіткій логіці розповсюдження отримали функції належності: трикутна, трапецеїдальна, гауссівська. Для даної задачі використаємо гауссівські функції належності – на рис. 2 показано всі терми обраної терм-множини {«слабкий зв'язок», «сильний зв'язок», «дуже сильний зв'язок»}.

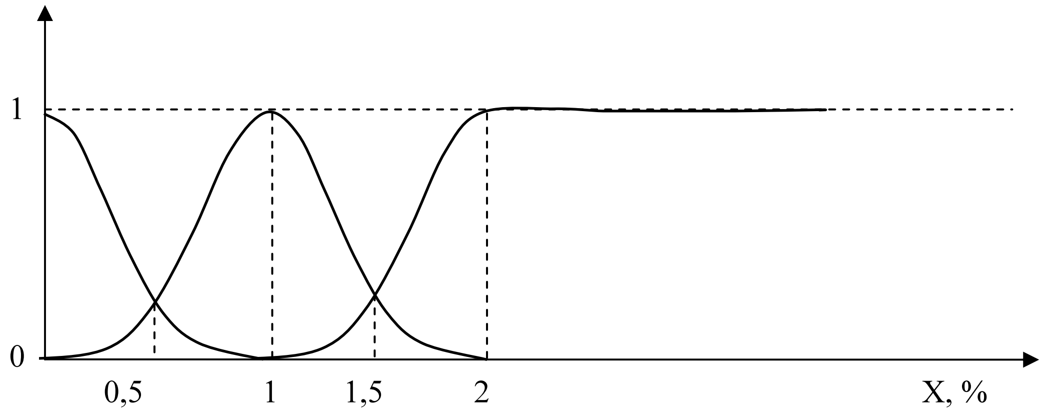



Рисунок 2 – Функції належності

Запропоновано вважати зв'язки, які зустрічаються менше 0,5 відсотка, слабкими, а ті, які зустрічаються частіше ніж 2% від усіх зв'язків – дуже сильними. Зрозуміло, що внаслідок масштабного аналізу багатьох різних текстів можна буде точніше визначити параметри та потрібну кількість функцій належності. Наприклад, якщо в обраному тексті розглядається якесь одне поняття, то це слово чи словосполучення буде зустрічатись дуже часто, а інші – набагато рідше, у такому випадку функція належності для терму «дуже сильний зв'язок» буде значно зсунута вправо і відокремлена від інших функцій. А для тексту, в якому фігурують декілька основних понять, функції належності для термів «сильний зв'язок» та «дуже сильний зв'язок» стануть вужчі і будуть відокремлені від функції належності для терму «слабкий зв'язок», оскільки багато понять будуть зустрічатись часто.

Запропонований у статті метод автоматизованої побудови онтологій дає непогані результати для української мови. Сучасні методи створення онтологій працюють

добре з англомовними текстами, проте для флексійної української мови все набагато складніше, оскільки порядок слів у реченні може бути довільним. На основі визначення синтаксичних зв'язків у реченні та побудови таблиці цих зв'язків, експериментально було підтверджено високу ефективність методу. Так, розглядаючи лише частину таблиці вищенаведеного прикладу розміром 20х20, що являє собою 2,07% від усієї таблиці, було отримано 85,7% всіх вагомих зв'язків у тексті.

Внаслідок дослідження обґрунтовано підхід до побудови моделі образної індексації електронного контенту, визначено основні задачі та особливості методу автоматизованої побудови онтологій. Використання нечіткої логіки у запропонованому методі забезпечує зручну експертну перевірку кінцевих результатів онтології у вигляді сили асоціативних зв'язків між мовними образами. Досягнення в експерименті 85,7% ефективності методу для української мови демонструє актуальність побудови парсеру для автоматизованого визначення синтаксичних зв'язків в україномовному реченні.

## Література

1. Валькман Ю.Р. Образы и образное мышление: некоторые отношения и структуры [Електронний ресурс] / Ю.Р. Валькман // V Междунар. науч.-практ. конф. «Интегрированные модели и мягкие вычисления в искусственном интеллекте», (Коломна, 28 – 30 мая 2009 г.). – Режим доступу : http://raai.org/resurs/papers/kolomna2009/doklad/Valkman.doc.
2. Бісікало О.В. Концептуальне поєднання понять образного мислення та мовленнєвої діяльності / О.В. Бісікало // Інформаційні технології та комп'ютерна інженерія. – 2010. – № 1 (17). – С. 72-77.
3. Бісікало О.В. Побудова нечітких відношення і простору сенсу образних конструкцій / О.В. Бісікало // Вісник Київського національного університету імені Тараса Шевченка. Серія: фізико-математичні науки. – 2011. – Вип. № 1. – С. 70-73.
4. Коваль А.А. Логіко-лінгвістичні моделі в нечітких системах / А.А. Коваль // Проблеми програмування. – 2008. – № 2-3. – С. 375-378.

## Literatura

1. Valkman Y.R. Images and image thinking: some relations and structures [Electronic source] / Y.R. Valkman // V International scientific-practical conference «Integrated models and soft computing in artificial intelligence», (Kolomna, May 28 – 30, 2009). – Available : http://raai.org/resurs/papers/kolomna2009/doklad/Valkman.doc.
2. Bisikalo O.V. Conceptual combination of notions of image thinking and speech activities / O.V. Bisikalo // Information Technology and Computer Engineering. – 2010. – № 1 (17). – P. 72-77.
3. Bisikalo O.V. Building of fuzzy sense relations and space of image constructions / O.V. Bisikalo // Announcer of Taras Shevchenko national university of Kyiv. Series: physical-mathematical sciences. – 2011. – Publ. № 1. – P. 70-73.
4. Koval A.A. Logical-linguistic models in fuzzy systems / A.A. Koval // Programming problems. – 2008. – № 2-3. – P. 375-378.

**_RESUME_**

**_O.V. Bisikalo, H.O. Kyrylenko_**

## *Using of Fuzzy Logic for Determining of Connection Strength Between Language Images*

Method of automated ontologies building is developed in this work, the method is based on image indexation of electronic content and uses fuzzy logic methods for determining of connection strength between language images.

Due to the research approach to building of model of electronic content image indexation is reasoned, main tasks and features of method of automated ontologies building are determined. Fuzzy logic using in the method provides easy expert verification of output ontology results. Experiment efficiency of 85,7% for Ukrainian language demonstrates parser building topicality for automated determining of syntactic connections in Ukrainian sentence.

*Стаття надійшла до редакції 26.04.2013.*